Context adaptive image denoising through modeling of curvelet domain statistics

Linda Tessens
Aleksandra Pižurica
Ghent University
TELIN-IPI-IBBT
St-Pietersnieuwstraat 41
B-9000 Ghent, Belgium
E-mail: Linda.Tessens@telin.ugent.be

Alin Alecu
Adrian Munteanu
Vrije Universiteit Brussel
ETRO-IBBT
Pleinlaan 2
B-1050 Brussels, Belgium

Wilfried Philips
Ghent University
TELIN-IPI-IBBT
St-Pietersnieuwstraat 41
B-9000 Ghent, Belgium

Abstract. *We perform a statistical analysis of curvelet coefficients, distinguishing between two classes of coefficients: those that contain a significant noise-free component, which we call the "signal of interest," and those that do not. By investigating the marginal statistics, we develop a prior model for curvelet coefficients. The analysis of the joint intra- and inter-band statistics enables us to develop an appropriate local spatial activity indicator for curvelets. Finally, based on our findings, we present a novel denoising method, inspired by a recent wavelet domain method called ProbShrink. The new method outperforms its wavelet-based counterpart and produces results that are close to those of state-of-the-art denoisers.*
 [DOI: 10.1117/1.2987723]

1 Introduction

In recent years, many novel geometric image transforms have been developed such as the ridgelet transform,[1] the wedgelet transform,[2] and the contourlet transform.[3] These new transforms capture the geometric information present in images, and in this sense overcome the limitations of classical wavelets. Among these, a mathematically elegant method entitled the "curvelet transform" has gained increasing popularity. Curvelets are directional basis functions that are highly localized, both in space and frequency. In Ref. 4, two digital implementations of the second-generation curvelet transform were presented. We refer the reader to this work for a comprehensive description of this transform.

The potential these geometric transforms hold for denoising images has been investigated by many researchers, e.g., in Refs. 5–8. As is the case when denoising images using the classical wavelet transform, noise reduction in the new transform domains results from greatly reducing the magnitude of the coefficients that contain primarily noise, while reducing others to a lesser extent. Thresholding as it has been applied in the wavelet domain[9] also has been used successfully in the curvelet and the contourlet domains.[6,8] Optimizing the choice of the threshold between these two classes of coefficients will improve the denoising performance of a method for wavelets[10,11] as well as for other transforms.[8] A very broad class of wavelet-based denoisers estimates the noise-free coefficients by minimizing a Bayesian risk, either by minimum mean-squared error or maximum *a posteriori* estimation.[12–15] These methods are optimized with respect to the marginal statistics of the coefficients within each sub-band by imposing a prior distribution on the noise-free transform coefficients. A particular success was exhibited by denoising methods where the local context was considered in the choice of one or more parameters of the prior model.[16–22] Recently, Po *et al.* transferred this reasoning to the contourlet domain.[7] To enable this transfer, marginal and joint image statistics on oriented multiscale pyramids, of which curvelets are a special case,

Paper 07187RR received Sep. 14, 2007; revised manuscript received Jun. 5, 2008; accepted for publication Jul. 6, 2008; published online Sep. 29, 2008. This paper is a revision of a paper presented at the SPIE conference on Wavelet Applications in Industrial Processing IV, October 2006, Boston, Massachusetts. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 6383.

1017-9909/2008/17(3)/033021/17/$25.00 © 2008 SPIE and IS&T.

were previously studied by Po *et al.* for the contourlet case, and by Boubchir *et al.* and Alecu *et al.* for the curvelet case.[7,23–25] In addition, Boubchir *et al.* proposed a multivariate prior model for curvelet coefficients.[23,24]

Over the past few years, mixture priors have been shown to be effective in wavelet processing.[11,13,22,26–29] Recently, we investigated how to develop a related prior for curvelet coefficients,[30] particularly one similar to that proposed by the authors in Ref. 22. As part of this endeavor, we extended the statistical analyses in Refs. 7 and 23–25 by investigating the different behaviors of curvelet coefficients that contain a significant noise-free component, and coefficients in which such a "signal of interest" is absent.[30,31] Based on our findings and inspired by the wavelet domain ProbShrink estimator,[22] we also defined and analyzed different intra-band[30] and inter-band[31] local spatial activity indicators (LSAIs) in the curvelet domain.

In this paper, we extend our previous statistical studies in a more comprehensive way, and introduce and analyze a new LSAI that includes both intra- and inter-band dependent curvelet coefficients. Using this new LSAI in the curvelet-based denoising method ProbShrinkCurv that we developed in Ref. 30 allows us to improve upon our previous denoising results reported in Refs. 30 and 31.

The remainder of this paper is organized as follows: Sec. 2 provides some background on the curvelet transform and introduces the notations and terminology used in this paper. Section 3 presents a comparative statistical analysis of the two classes of curvelet coefficients mentioned above. Section 4 discusses a novel curvelet-based context adaptive denoising method, and Sec. 5 considers the parameter that marks the threshold between the coefficient classes. Sections 6 and 7 summarize the main results and conclusions of this paper.

2 Curvelet Transform

Although the curvelet transform was originally a continuous transform,[32] it has several digital approximations. The two most recent ones were introduced by Candès *et al.*[4] There, one implementation is based on unequally spaced fast Fourier transforms (USFFT), while the other is based on the wrapping of specially selected Fourier samples.[4] Since the latter implementation exhibits a somewhat faster running time, especially for the inverse transform (see Ref. 4), we use it throughout this paper.

The curvelet transform decomposes the image in several frequency bands. At the coarsest scale, isotropic wavelets are used as basis functions. At finer scales, curvelets take over this role. At the finest scale (highest frequencies), one can choose between a wavelet and a curvelet decomposition. Using curvelets at all scales leads to a transform that provides approximate rotation invariance (sharp directional selectivity), which is beneficial for denoising applications. This is why in this work we always choose a curvelet decomposition at the finest scale, in spite of disadvantages such as some possible aliasing and an increased redundancy of the transform (e.g., for a 512×512 image, with 4 scales in the curvelet decomposition and 16 orientations at the coarsest curvelet level, redundancy increases from 2.74 to 7.16 with curvelets at the finest scale).

Figure 1(b) shows the curvelet decomposition of the test image in Fig. 1(a) into 4 frequency scales with 8 orientations at the coarsest curvelet scale. The low-pass image is located at the center of the representation. The curvelet coefficients are arranged around it. For representation purposes, we display the magnitudes of the coefficients. Those with value zero are marked in white, whereas coefficients with large magnitudes are dark. From the prevalent white color of Fig. 1(b), it is clear that the curvelet decomposition of this image is extremely sparse. In fact, curvelets are known to lead to an optimally sparse representation of piecewise smooth images with discontinuities along smooth edges, and this with the best M-term nonlinear approximation.[4] The approximation error decays as $O[(\log M)^3 M^{-2}]$; hence, by the Bernstein inequality, the observed decay of the absolute values of the curvelet coefficients is known to be very fast.

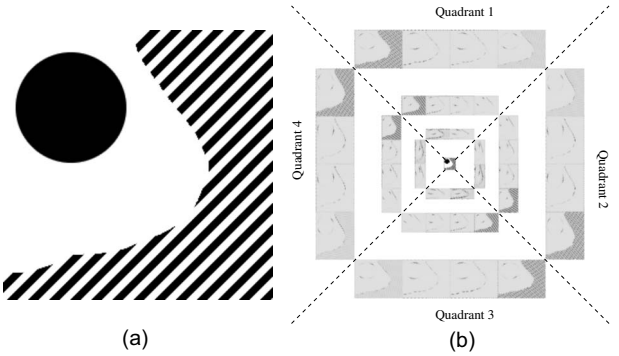

Fig. 1 (a) A 256×256 test image. (b) Curvelet decomposition of the test image into 4 scales and with 8 orientations at the coarsest scale. The low-pass image is located at the center of the representation. Curvelet coefficients with value zero are marked in white, whereas coefficients with a large magnitude are dark. The dotted lines mark the border between the four quadrants.

The curvelet coefficients are grouped according to orientation and scale. The concentric coronae represent the different scales starting with the lowest scale (low frequencies) in the center. Sub-bands of the same scale are ordered within these coronae so that the orientation suggested by their position matches the spatial frequencies they represent. Thus, the diagonal lines in Fig. 1(a) produce high curvelet coefficients in the sub-bands along the direction perpendicular to them. One can clearly discern four quadrants in each corona, which are numbered in a clockwise direction starting with the upper quadrant. Quadrant 1 contains the magnitudes of the real parts, and quadrant 3 contains the magnitudes of the imaginary parts of the (complex) curvelet coefficients produced by mostly horizontally oriented curvelet functions. Mutatis mutandis, the same holds for quadrants 2 and 4. So although the coefficients contained in the different quadrants are actually only the real or imaginary parts of the (in reality) complex curvelet coefficients, we refer to these real and imaginary parts as "curvelet coefficients."

2.1 *Terminology and Notations*

In this paper we use the same terminology as Po *et al.*,[7] Boubchir *et al.*,[23] and Alecu *et al.*[25] Given a curvelet coefficient X, we will use the following notations:

- N_i denotes the neighboring curvelet coefficient in the

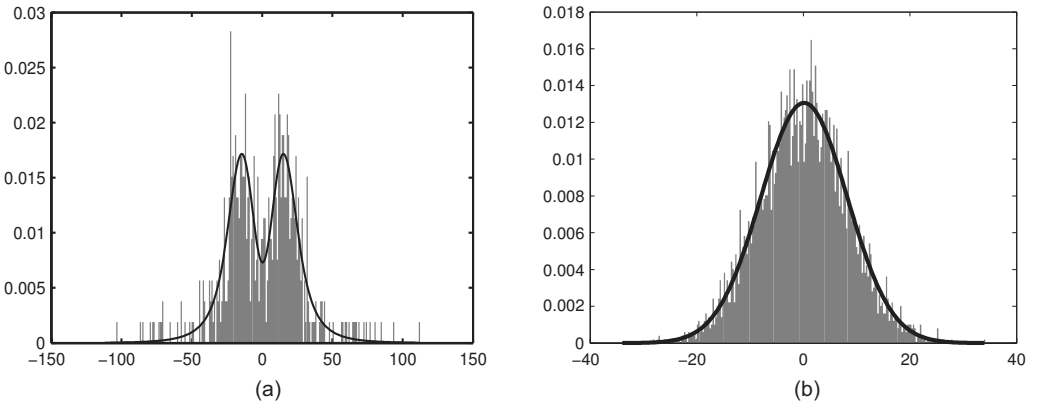

Fig. 2 Histograms of noisy curvelet coefficients (a) with noise-free component larger than a threshold T, and (b) in which no signal of interest is present. Overlayed on the histograms, the estimated pdf's of the significant and insignificant noise-contaminated curvelet coefficients are $f(x|H^{0,1})*\phi(0,\sigma')$.

same sub-band (all neighbors are numbered from 1 to 8, starting with the neighbor located at the upper left and proceeding clockwise).

- C_k denotes the cousin curvelet coefficient located at the same relative position as X in a different sub-band at the same scale, and k denotes the sub-band (all sub-bands within each corona are numbered, starting from 1 with the sub-band at the upper left and proceeding clockwise).
  - C_k is an adjacent cousin of X if sub-band k lies next to the sub-band in which X is located.
  - C_k is an opposing cousin of X if sub-band k lies opposite the sub-band in which X is located, or in other words, if C_k is the real (imaginary) part of the complex curvelet coefficient of which X is the imaginary (real) part.
- P denotes the parent curvelet coefficient located at the same relative position as X in the same sub-band but at a coarser scale.

3 Curvelet Statistics

Image statistics in the curvelet domain were recently studied by Boubchir *et al.*[23,24] and Alecu *et al.*[25] with a comparative analysis to wavelet domain statistics. Po *et al.*[7] did the same for the contourlet transform. We will now take a step further in this direction by analyzing the statistics of two classes of curvelet coefficients: those containing a significant noise-free component (which we will refer to as "significant" coefficients), and the coefficients in which no signal of interest is present (which we will call "insignificant"). For this statistical analysis we will make use of both noise-free and noisy image versions. The use of noise-free image versions enables the development of contextual models that we employ in the actual denoising procedure presented in Sec. 4, where noise-free image versions are not available.

In our approach, significant coefficients are defined as those that have a noise-free component larger in absolute value than a threshold T. We call such a component our signal of interest. In our statistical analysis experiments, we determine the locations of the significant noisy curvelet coefficients by thresholding their noise-free counterparts. Hence, a noisy curvelet coefficient is marked as significant if the corresponding curvelet coefficient of the noise-free image version exceeds a threshold T in magnitude.

An important issue at this point is the choice of the parameter T. This choice cannot be considered independently from the goal of this paper: developing a denoising method for curvelets that aims to minimize the mean squared error between the denoised and the noise-free image (the method will be described in detail in Sec. 4). Therefore, we postpone a discussion of this parameter to Sec. 5. Let it suffice for now to say that the threshold will be related to the standard deviation of the noise through a constant factor.

All the statistics in this section were obtained from the curvelet decompositions of images contaminated with additive white Gaussian noise (AWGN) with standard deviation $\sigma=20$ unless explicitly mentioned otherwise. The threshold between significant and insignificant coefficients was set to 1.3σ.

3.1 *Marginal Statistics*

Figure 2(a) shows the 256-bin histogram of the significant curvelet coefficients of a sub-band at the finest scale of the curvelet decomposition obtained from a noisy version of the image *Peppers*. Figure 2(b) shows the same histogram for the insignificant curvelet coefficients of this sub-band.

In Refs. 23–25, the authors showed that the probability density functions (pdfs) of noise-free curvelet coefficients x follow well a generalized Laplacian (also called generalized Gaussian) distribution. In the following, we denote the pdf of x by $f(x)$, so $f(x)=\upsilon/2s\Gamma(1/\upsilon)\exp(-|x/s|^{\upsilon})$, where s and υ are parameters of the generalized Laplacian distribution.

We adopt the modeling framework proposed by Pižurica and Philips[22] for significant and insignificant wavelet coefficients, and we apply it in the curvelet domain as follows. Let H^1 denote the hypothesis that a curvelet coefficient x is significant, and let H^0 denote the opposite. By our defini-

tion of significant curvelet coefficients, the pdf of these coefficients can be modeled by the tails of a generalized Laplacian:

$$f(x|H^1) = Af(x) \text{ for} |x| > T \text{ and } f(x|H^1) = 0 \text{ otherwise}, \quad (1)$$

where A is a normalizing constant. Similarly, the pdf of noise-free insignificant coefficients has the shape of the central part of a generalized Laplacian, or

$$f(x|H^0) = Bf(x) \text{ for} |x| \leqslant T \text{ and } f(x|H^0) = 0 \text{ otherwise}, \quad (2)$$

where B is a normalizing constant.

Now we investigate the distributions of the noisy curvelet coefficients when the input noise is AWGN with standard deviation σ. The curvelet transform, which is a linear transform, transforms AWGN into additive correlated Gaussian noise in each sub-band. The first-order pdf of the noise in the sub-bands is a normal distribution $\phi(0,\sigma')$. Since the second-generation curvelet transform we are using corresponds to a tight frame, the standard deviation σ' is $\sigma/\sqrt{\alpha}$, with α the redundancy factor of the transform. The pdf's of the significant and insignificant noisy curvelet coefficients can be modeled by the distribution of their respective noise-free counterparts, convolved with this normal distribution $\phi(0,\sigma')$: $f(x|H^{0,1}) * \phi(0,\sigma')$.[†] We estimate $f(x)$, which is necessary for the calculation of $f(x|H^1)$ and $f(x|H^0)$, from the noise-contaminated curvelet coefficients by using the method of Simoncelli *et al.*[14] for estimating the generalized Laplacian distribution from wavelet coefficients contaminated with AWGN. Although, as pointed out earlier, the noise in the curvelet sub-bands is not white, Fig. 2 shows that this approximate model matches well with observations.

The proposed prior model belongs to a broader class of finite mixtures of two distributions, one modeling the statistics of significant coefficients and the other one of the insignificant coefficients.[11–13,26–29,33] In earlier models of this class, the mixed distributions are usually two normal distributions (e.g., in Ref. 12), a normal distribution and a point mass at zero (e.g., in Refs. 11 and 13), or a Laplacian distribution and a point mass at zero.[33] For such prior models, the mixing proportion [i.e., $P(H^0)$] as well as the hyperparameters are usually estimated jointly using a maximum likelihood (ML) estimator with an expectation-maximization (EM) algorithm. On the contrary, the proposed prior has the double advantage that the parameter estimation procedure is simpler (no iterative joint estimation necessary) and that it can cope with a more complicated model of the noise-free coefficients (generalized Laplacian). The only parameter that cannot be directly estimated from the data is the threshold T. The choice of this parameter is discussed in Sec. 5.

3.2 *Joint Statistics*

Previous studies revealed that noise-free curvelet coefficients are strongly correlated in local intra-band neighborhoods, and that these local correlations are stronger than their inter-scale and inter-orientation counterparts.[23–25] We will now investigate whether the joint statistics of significant coefficients differ from the joint statistics of coefficients in which no signal of interest is present. Such a difference would facilitate the denoising of the coefficients. In particular, we will focus on the magnitude of the coefficients.

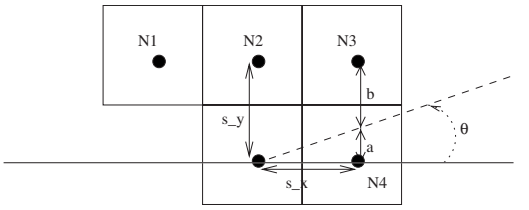

Fig. 3 A curvelet coefficient and four of its neighbors (labeled N1 to N4). The curvelet basis function is oriented in the direction θ.

The correlation coefficients in this section have been calculated as the average of the correlation coefficients obtained from the magnitudes of the next-to-highest frequency scale curvelet coefficients of a test set of 44 images.[‡] As test images, we chose the images from the Miscellaneous volume of the USC-SIPI image database (converted to grayscale).[34] To better evaluate the general validity of the results, we also report the correlation coefficients for four specific images: *House, Peppers, Barbara*, and *Baboon*.

3.2.1 *Intra-band correlations*

In Refs. 23–25, it was shown that each curvelet coefficient is strongly correlated with its eight direct neighbors. Alecu *et al.* additionally showed that the correlation is not equally strong for all the neighbors. By construction, curvelet coefficients are more correlated with the neighbors that lie in the direction parallel to the main direction of the curvelet function by which they were produced. In this section, we examine whether the same holds for the two categories of coefficients that are of interest to us: the significant and the insignificant ones.

In Ref. 25, Alecu performed his analysis for the USFFT implementation of the curvelet transform (see Ref. 4). For this implementation, the main direction of the curvelet function is normalized to be either horizontal for the sub-bands in quadrants 1 and 3 or vertical for the sub-bands in quadrants 2 and 4, where quadrants are numbered as in Fig. 1(b). This means that the direction of highest correlation always coincides with one of the grid lines on the discrete image raster. This is not the case with the wrapping implementation, which we use throughout this paper. Instead, the curvelet basis functions have an orientation that differs from sub-band to sub-band and can be characterized by an angle θ with respect to the horizontal direction. Correlation is maximal along this direction. To investigate the correlation properties along it, interpolation of neighboring coefficients is required.

[†]The convolutions in these expressions and in the remainder of this paper are numerically evaluated by multiplying the FFTs of the two separate sequences and then calculating the inverse FFT.

[‡]If the insignificant coefficients were pure noise, their correlation coefficients could also be obtained by calculating the sample covariance matrix of the curvelet decomposition of a scaled delta function (which has the same power spectrum as AWGN).[20] However, although the relative influence of the noise is much bigger on the insignificant coefficients than on the significant ones, insignificant coefficients are not exactly the same as noise, and therefore we will not adopt the theoretical method described above for the computation of their correlation coefficients.

Table 1 Mean correlation coefficients between significant and insignificant curvelet coefficient magnitudes and (a) neighbors in the direction of highest correlation, (b) neighbors in the direction of lowest correlation, (c) adjacent cousins, (d) opposing cousins, and (e) parents.

Coefficients conditioned on:	Correlation coefficients			
	Significant coefficients		Insignificant coefficients	
	Mean	Std. Dev.	Mean	Std. Dev.
(a) Neighbors correlated	0.46	0.06	0.26	0.04
(b) Neighbors uncorrelated	0.10	0.11	0.03	0.02
(c) Adjacent cousins	0.16	0.08	0.03	0.02
(d) Opposing cousins	0.14	0.07	0.02	0.02
(e) Parents	0.17	0.07	0.04	0.02

Consider, for example, a curvelet coefficient and its four neighbors arranged as in Fig. 3. Since the direction of highest correlation, characterized by θ, does not coincide with one of the grid lines, none of the four neighbors lies exactly in the direction of highest correlation. N_3 and N_4 are closest, so we linearly interpolate them:

$$N' = \frac{a}{a+b}|N_3| + \frac{b}{a+b}|N_4|, \tag{3}$$

where $a = s_x \tan\theta$, $b = s_y - s_x \tan\theta$, and s_x is the horizontal and s_y the vertical sampling period. For other values of θ, N_3 and N_4 must be substituted by the appropriate neighbors, and the interpolation weights a and b should be adapted accordingly.

In the following we will always analyze the correlation between a reference coefficient and this coefficient N', calculated from its neighbors. We will refer to N' as the neighbor lying in the direction of highest correlation. Of course, by symmetry, each coefficient has two such neighbors. A similar reasoning applies to the analysis of correlation in the direction perpendicular to the direction of highest correlation, i.e., the direction of smallest correlation.

Table 1 (a) lists the mean correlation coefficients between the significant and insignificant coefficient magnitudes and one of their two neighbors in the direction of highest correlation, calculated from the next-to-highest-scale sub-bands of our image test set. These large mean correlation coefficients match with theoretical expectations of high correlation in the direction of the curvelet basis function. In the perpendicular direction of smallest correlation, one observes that this correlation virtually disappeared for the insignificant coefficients and became very small for the significant ones [see Table 1 (b)]. For the latter class of coefficients, the standard deviation of the correlation coefficients calculated over the image test set is of approximately the same magnitude as the average value. This means that correlation with this kind of coefficients is highly image-dependent. Indeed, correlation is quite high for *Barbara* and *Baboon* but almost disappeared for *Peppers* and *House* [see Table 2 (b)]. Unlike *Barbara* and *Baboon*, *Peppers* and *House* both correspond better to the image model for which curvelets are especially suited, namely piecewise smooth with discontinuities along the curvilinear edges. Indeed, *Barbara* and *Baboon* both have a less sparse curvelet representation than *Peppers* and *House*. In a decomposition into four scales and with 16 orientations at the coarsest level, 11.81% of the coefficients are classified as significant for *Baboon* and 5.08% for *Barbara* versus only 2.62% for *Peppers* and 3.49% for *House*. Thus, the achieved decorrelation of the transform coefficients is higher for these two images.

Table 2 Correlation coefficients for some specific images between significant and insignificant curvelet coefficient magnitudes and (a) neighbors in the direction of highest correlation, (b) neighbors in the direction of lowest correlation, (c) adjacent cousins, (d) opposing cousins, and (e) parents.

Coefficients conditioned on:	Correlation coefficients							
	Significant coefficients				Insignificant coefficients			
	House	Peppers	Barbara	Baboon	House	Peppers	Barbara	Baboon
(a) Neighbors correlated	0.43	0.50	0.51	0.45	0.28	0.28	0.27	0.23
(b) Neighbors uncorrelated	0.07	0.08	0.35	0.19	0.02	0.02	0.03	0.03
(c) Adjacent cousins	0.11	0.19	0.11	0.18	0.02	0.03	0.04	0.03
(d) Opposing cousins	0.08	0.19	0.22	0.19	0.02	0.02	0.02	0.03
(e) Parents	0.22	0.25	0.04	0.17	0.03	0.04	0.03	0.03

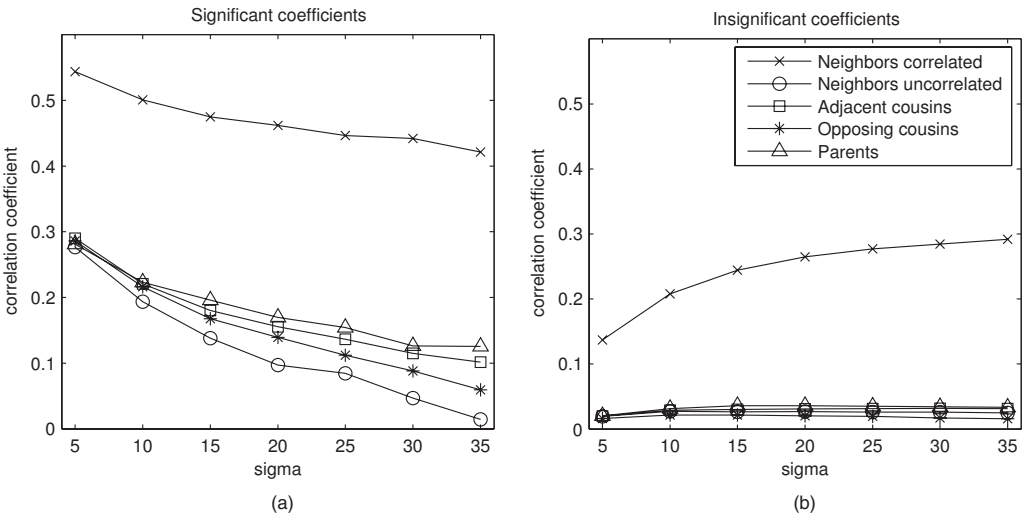

Fig. 4 Correlation coefficients between (a) significant coefficients, and (b) insignificant coefficients and their neighbors in the direction of highest and lowest correlation, adjacent and opposing cousins and parents, as a function of the standard deviation σ of the contaminating AWGN.

As mentioned at the start of Sec. 3, the images from the test set were contaminated with AWGN (σ=20) to obtain the correlation coefficients in Table 1. Figure 4 shows the evolution of the correlation coefficients between significant coefficients and their neighbors and insignificant coefficients and their neighbors in the direction of highest and lowest correlation as a function of the standard deviation σ of the AWGN with which the images in the test set were contaminated. To avoid overloading the graph, error bars were omitted. The authors verified that the standard deviations of the correlation coefficients were approximately constant for all values of σ; for σ=20, these numbers can be found in Table 1. For significant coefficients, the relative relationship of the two curves is maintained for all values of σ. For insignificant coefficients, correlation with neighbors in the direction of lowest correlation remains negligible regardless of σ, whereas the correlation with neighbors in the direction of highest correlation displays a rising trend.

3.2.2 *Inter-band dependencies*

Boubchir *et al.*[23,24] and Alecu, *et al.*[25] found that curvelet coefficients of different sub-bands are approximately decorrelated, but some dependencies between sub-bands do exist. Boubchir *et al.* and Alecu *et al.* observed dependency between a curvelet coefficient and its parent, as well as between a curvelet coefficient and its cousins. The strength of these inter-orientation dependencies decreases with an increase in the difference of orientation, but dependency is also observed with respect to the opposite orientation.[25] We will now further extend this study of inter-band curvelet statistics to the two categories of coefficients that we consider in this paper: the significant and the insignificant curvelet coefficients.

The average correlation coefficients between significant, respectively insignificant coefficient magnitudes and the magnitudes of their adjacent and opposing cousins and their parents are indicated in Table 1 (c) to (e). The same observations can be made here as in the previous section for the correlation coefficients in the direction of smallest correlation: correlation has virtually disappeared for the insignificant coefficients and is small for the significant ones. Figure 4 confirms that this observation also holds for other standard deviation values of the contaminating AWGN. The high standard deviation of the correlation coefficients for the significant coefficients implies that the correlation is highly image-dependent [see Table 1 (c) to (e)]. This is confirmed by the values in Table 2 (c) to (e).

3.3 *Local Spatial Activity Indicators*

Inspired by the ProbShrink wavelet domain denoising method of Ref. 22 that was developed by some of the authors of this paper, we now define and analyze different LSAIs in the curvelet domain. In general, for each curvelet coefficient we define the LSAI as a function of those coefficients that are well correlated when the coefficient is significant. We investigated this last property in the previous section about joint curvelet statistics (Sec. 3.2), and this study led us to propose four LSAIs that will potentially perform well in the curvelet denoiser that will be developed in Sec. 4: two intra-band,[30] one inter-band,[31] and one novel combined intra-inter-band (IIB) LSAI. We will evaluate these LSAIs, first by investigating the correlation between the magnitudes of the curvelet coefficients and their corresponding LSAI, since this is the correlation that is exploited in our newly developed denoising method (see Sec. 4), and second, by plugging them into this denoising method. The second evaluation will be performed in the next section.

As in the previous section, the correlation coefficients in this section have been calculated as the average correlation coefficients obtained from the next-to-highest-frequency scale sub-bands of the USC-SIPI image test set, and results are also reported for four specific images (*House*, *Peppers*, *Barbara*, and *Baboon*).

3.3.1 *Anisotropic intra-band LSAIs*

Because of the correlation properties of curvelet coefficients, an anisotropic LSAI seems appropriate. We propose

Table 3 Average correlation coefficients between the magnitudes of significant and insignificant curvelet coefficients and (a) the anisotropic LSAI oriented in the direction of highest correlation, (b) the anisotropic LSAI oriented in the direction of lowest correlation, (c) the adjacent, opposing, and parents LSAI, and (d) the combined IIB LSAI.

	Correlation coefficients			
	Significant coefficients		Insignificant coefficients	
LSAI	Mean	Std. Dev.	Mean	Std. Dev.
(a) Anisotropic LSAI correlated	0.62	0.06	0.33	0.06
(b) Anisotropic LSAI uncorrelated	0.12	0.13	0.04	0.03
(c) AOP LSAI	0.26	0.09	0.06	0.02
(d) Combined IIB LSAI	0.20	0.12	0.05	0.03

two anisotropic LSAI candidates z that are calculated as the mean absolute value of the $n-1$ coefficients N' within a small $1\times n$ neighborhood δ (that excludes the reference coefficient itself) oriented in the direction of either highest or lowest correlation:

$$z=\frac{1}{n-1}\sum_{i\in\delta_{(\mathrm{un})\mathrm{corr}}}|N_i'|. \tag{4}$$

The coefficients N_i' within this window δ are interpolated from their neighbors, as explained in Eq. (3)

Table 3 (a) and (b) show the average correlation coefficients between the curvelet coefficient magnitudes of the next-to-finest-scale sub-bands of the images from the test set and the LSAI oriented in the direction of highest and lowest correlation (in both cases, n is set to 5).

The correlation coefficients for these LSAIs follow the trend of those between a coefficient and its neighbor, discussed in Sec. 3.2.1 [see Table 1 (a) and (b)]. Indeed, the anisotropic LSAI oriented in the direction of highest correlation is highly correlated with both the significant and the insignificant coefficients. Again, this correlation is higher for the significant coefficients than for the insignificant ones. For both classes, this correlation is also larger than the correlation between a coefficient and just one of its neighbors, indicated in Table 1 (a), since the LSAI summarizes information from more coefficients (here, 4 as opposed to only 1) and can also capture correlations over longer distances. For the anisotropic LSAI oriented in the direction of lowest correlation, correlation is low in both cases, although slightly higher for the significant coefficients. Again, the standard deviation of the correlation coefficients is very high in the significant case, which means that this correlation is highly image-dependent (see Table 4 for correlation coefficients of some specific images). Again, we observe that the decorrelation of the curvelet coefficients in the direction of smallest correlation is highest for *Peppers* and *House*, both sparsely represented in the curvelet domain.

3.3.2 *Adjacent, opposing, and parents (AOP) inter-band LSAI*

In previous work,[31] we defined and discussed several inter-band LSAIs. Different LSAIs were calculated for each curvelet coefficient: the average magnitude of the adjacent cousins; the adjacent and opposing cousins; the adjacent cousins and the parent; or the adjacent and opposing cousins and the parent (AOP). The last LSAI proved to be the best performing one in terms of denoising capabilities. Therefore, we will discuss only this inter-band LSAI here.

For a coefficient y in a sub-band k, the LSAI is defined as

$$z=\frac{1}{4}(|C_{k-1}|+|C_{(k+1)\bmod \mathrm{K}}|+|C_{(k+K/2)\bmod \mathrm{K}}|+|P|),$$
$$k\in\{1,\ldots,K\},\ \text{and}\ \mathrm{C}_0=\mathrm{C}_\mathrm{K}, \tag{5}$$

where "mod" stands for the modulo operation, K is the

Table 4 Correlation coefficients for some specific images between the magnitudes of significant and insignificant curvelet coefficients and (a) the anisotropic LSAI oriented in the direction of highest correlation, (b) the anisotropic LSAI oriented in the direction of lowest correlation, (c) the adjacent, opposing, and parents LSAI, and (d) the combined IIB LSAI.

	Correlation coefficients							
	Significant coefficients				Insignificant coefficients			
Coefficients conditioned on:	House	Peppers	Babara	Baboon	House	Peppers	Babara	Baboon
(a) Anisotropic LSAI corr.	0.58	0.67	0.65	0.60	0.35	0.37	0.33	0.27
(b) Anisotropic LSAI uncorr.	0.11	0.07	0.42	0.27	0.03	0.04	0.05	0.04
(c) AOP LSAI	0.27	0.35	0.25	0.30	0.05	0.06	0.07	0.06
(d) Combined IIB LSAI	0.20	0.21	0.44	0.31	0.04	0.05	0.06	0.05

number of orientations at the scale to which y belongs, and the notations introduced in Sec. 2.1 are used.

In Table 3 (c), the average correlation coefficients between the curvelet coefficient magnitudes and the AOP LSAI are indicated. The insignificant curvelet coefficients are approximately decorrelated with this LSAI, whereas some correlation exists between the significant coefficients and this LSAI. Table 4 (c) shows that this behavior is present for all our example images.

3.3.3 *Combined intra- and inter-band LSAI*

To exploit both the intra- and inter-band correlations between curvelet coefficients, we now define a novel LSAI that combines the best performing intra- and inter-band LSAIs. Specifically, we define an IIB LSAI as the average of the anisotropic intra-band LSAI oriented in the direction of lowest correlation and the AOP inter-band LSAI:

$$z=\frac{1}{2}\left[\frac{1}{4}(|C_{k-1}|+|C_{k+1 \bmod \mathrm{K}}|+|C_{k+K/2 \bmod \mathrm{K}}|+|P|)+\frac{1}{n-1}\sum_{i\in\delta_{\mathrm{uncorr}}}|N_i'|\right],\quad k\in\{1,\ldots,K\},\ \text{and}\ \mathrm{C}_0=\mathrm{C}_{\mathrm{K}}. \tag{6}$$

Table 3 (d) indicates the average correlation coefficients over our image test set between the curvelet coefficient magnitudes and this candidate LSAI. Table 3 (d) again shows that the insignificant curvelet coefficients are approximately decorrelated with this LSAI. The average correlation coefficient for the significant coefficients is lower than in the AOP LSAI case.

4 Context-Adaptive Image Denoising Using Curvelets

Based on our findings in Secs. 3.1 and 3.3, we now develop a curvelet domain version ProbShrinkCurv of the ProbShrink denoising method.[22]

4.1 *ProbShrinkCurv Denoiser*

Consider an input image contaminated with AWGN. After transforming this image to the curvelet domain, let y_l denote, for a given scale and orientation, the curvelet coefficient at position l. Let y_l be composed of an unknown noise-free curvelet coefficient x_l and of a noise component n_l: $y_l=x_l+n_l$, where the variables n_l are identically distributed Gaussian random variables that are statistically independent from y_l. Let H_l^1 denote the hypothesis that x_l represents a significant image feature, and H_l^0 denote the hypothesis that x_l contains no signal of interest. The hypothesis H_l^1 is equivalent to $|x_l|>T$, and H_l^0 is equivalent to $|x_l|\leqslant T$, where T is an empirically determined threshold (see Sec. 5). Finally, let z_l be any arbitrary indicator of the local spatial activity defined as in Sec. 3.3.

The minimum mean squared error (MMSE) estimate of x_l is[35,36]

$$\hat{x}_l=E(x_l|y_l,z_l)=E(x_l|y_l,z_l,H_l^1)P(H_l^1|y_l,z_l)+E(x_l|y_l,z_l,H_l^0)P(H_l^0|y_l,z_l). \tag{7}$$

Since H_l^0 refers to the absence of a signal of interest, $E(x_l|y_l,z_l,H_l^0)=0$. We further assume we can approximate $E(x_l|y_l,z_l,H_l^1)$ by y_l. This leads to

$$\hat{x}_l=P(H_l^1|y_l,z_l)y_l. \tag{8}$$

Using Bayes' rule, we can rewrite this expression as

$$\hat{x}_l=\frac{\Lambda_l}{1+\Lambda_l}y_l, \tag{9}$$

where Λ_l is the general likelihood ratio $\Lambda_l=\rho_l\zeta_l\upsilon_l$, with $\rho_l=P(H_l^1)/P(H_l^0)$, $\zeta_l=p(z_l|H_l^1)/p(z_l|H_l^0)$ and $\upsilon_l=p(y_l|H_l^1)/p(y_l|H_l^0)$. Applying the inverse curvelet transform to the estimated noise-free curvelet coefficients $\hat{x}_l$ yields the denoised image.

We will now comment in detail on the calculation of each of the factors of Λ_l.

4.2 *Calculation of ρ_l*

The factor ρ_l can be rewritten as[22]

$$\rho_l=\frac{P(H_l^1)}{P(H_l^0)}=\frac{\int_{-\infty}^{+\infty}f(x_l|H^1)dx_l}{\int_{-\infty}^{+\infty}f(x_l|H^0)dx_l}=\frac{\int_{-\infty}^{-T}f(x_l)dx_l+\int_{T}^{\infty}f(x_l)dx_l}{\int_{-T}^{T}f(x_l)dx_l}=\frac{1-\int_{-T}^{T}f(x_l)dx_l}{\int_{-T}^{T}f(x_l)dx_l}. \tag{10}$$

As mentioned in Sec. 3.1, it was shown in Refs. 23–25 that the noise-free curvelet coefficients x follow well a generalized Laplacian. So with $f(x_l)=\upsilon/2s\Gamma(1/\upsilon)\exp(-|x_l/s|^{\upsilon})$, and by substituting $t=(x_l/s)^{\upsilon}$, then $\int_{-T}^{T}f(x_l)dx_l$ can be obtained as

$$\int_{-T}^{T}f(x_l)dx_l=\frac{\nu}{s\Gamma\left(\frac{1}{\nu}\right)}\int_0^T\exp\left(-\left(\frac{x_l}{s}\right)^{\nu}\right)dx_l=\frac{1}{\Gamma\left(\frac{1}{\nu}\right)}\int_0^{(T/s)^{\nu}}t^{1/\nu-1}e^{-t}dt=\Gamma_{\mathrm{inc}}\left(\left(\frac{T}{s}\right)^{\nu},\frac{1}{\nu}\right), \tag{11}$$

where $\Gamma_{\mathrm{inc}}(y,a)=1/\Gamma(a)\int_0^y t^{a-1}e^{-t}dt$ is the incomplete gamma function. Thus, ρ_l becomes

$$\rho_l=\frac{1-\Gamma_{\mathrm{inc}}\left[\left(\frac{T}{s}\right)^{\nu},\frac{1}{\nu}\right]}{\Gamma_{\mathrm{inc}}\left[\left(\frac{T}{s}\right)^{\nu},\frac{1}{\nu}\right]}. \tag{12}$$

4.3 *Calculation of υ_l*

For the computation of υ_l, we showed in Sec. 3.1 that $p(y_l|H_l^{0,1})$ can be modeled as $f(x|H^{0,1})*\phi(0,\sigma')$, with $f(x|H^{0,1})$ as defined in Eqs. (1) and (2).

4.4 *Calculation of ζ_l*

The calculation of ζ_l depends on the choice of the LSAI z_l. The choice of this LSAI will be based on the study of joint curvelet statistics in Sec. 3.3. We will discuss this choice

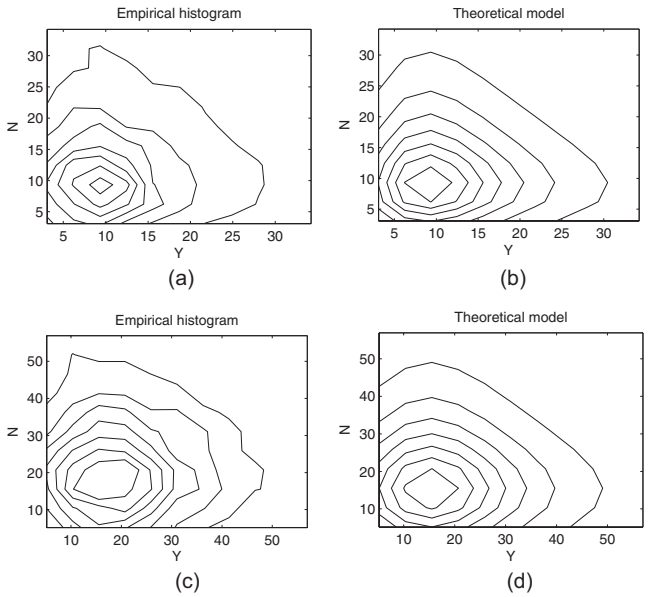

Fig. 5 Joint histogram of a significant coefficient and its significant neighbors within a 1×5 spatial neighborhood δ, oriented in the direction of lowest correlation, for a sub-band of the next-to-highest scale of (a) *Barbara* contaminated with AWGN ($\sigma=10$), (c) *Baboon* contaminated with AWGN ($\sigma=20$), and (b) and (d) approximation by a conditionally independent model $p(|y||H^1)*p(|y||H^1)$.

later in this section. First we will focus on the derivation of the pdf's of the LSAI, conditioned on either the hypothesis H_l^0 or H_l^1. For compactness of notation, we will suppress the position index l in what follows.

When z is an intra-band LSAI, i.e., $z=1/(n-1)\,\Sigma_{i\in\delta_{(\mathrm{un})\mathrm{corr}}}|N_i'|$, it is calculated from coefficients that lie within a small spatial neighborhood δ around the central coefficient. The statistical characterization of the LSAI is greatly simplified by assuming, as in Refs. 19 and 22, that all the coefficient magnitudes within this small neighborhood, including the ones that are interpolated from their neighbors, are identically distributed [either as $p(|y||H^1)$ or as $p(|y||H^0)$] and are conditionally independent (given H^0 or H^1). Under these assumptions, $p(z|H^{0,1})$ can be approximated by convolving $p(|y||H^{0,1})$ with itself $n-1$ times.

The first assumption of identical distribution is likely because neighboring coefficients are statistically dependent. To justify the second assumption of conditional independence, we experimentally show that neighbors are not much more correlated than being from the same class of pdf (significant or insignificant). To that end, Fig. 5(a) plots the joint histogram of the magnitudes of a significant coefficient and its significant neighbors within a 1×5 spatial neighborhood δ, oriented in the direction of lowest correlation, for a sub-band of the next-to-highest scale of *Barbara*, contaminated with AWGN ($\sigma=10$). The correlation coefficient between the magnitudes of significant coefficients and their neighbors in the direction of lowest correlation is 0.44 for this sub-band. Figure 5(b) shows the approximation by a conditionally independent model $p(|y||H^1)*p(|y||H^1)$. A comparison of Figures 5(a) and 5(b) reveals that the model matches well with the empirical histogram, despite the correlation between the coefficients. Figures 5(c) and 5(d) show the same for *Baboon* contaminated with AWGN ($\sigma=20$). For this sub-band, the correlation coefficient between significant coefficient magnitudes and their neighbors in the direction of lowest correlation is 0.07.

For the calculation of the pdf of the AOP inter-band LSAI, we make the same assumptions. More precisely, we assume that the coefficients from other sub-bands that are incorporated into the LSAI are identically distributed as the reference coefficient y_l, conditioned on either H^1 or H^0, and are conditionally independent (given H^0 or H^1). Thus, in this case $p(z|H^{0,1})$ can be approximated by convolving $p(|y||H^0)$ with itself 4 times.

The combined intra- and inter-band LSAI is computed as the average of the intra-band LSAI oriented in the direction of lowest correlation and the AOP LSAI, so its pdf can be obtained through the convolution of the pdf's of the LSAIs it is calculated from.

4.5 *Choice of the LSAI*

We will now evaluate the denoising potential of the LSAIs that were proposed and studied in Sec. 3.3 based on the results of Sec. 3.2. We used the ProbShrinkCurv method (4 scales in the curvelet decomposition, 16 orientations at the coarsest level) with the different LSAIs proposed in Sec. 3.3 to denoise some 512×512 grayscale images contaminated with AWGN with standard deviations of 5, 10, 20, 30, and 50. The peak signal-to-noise ratio (PSNR) results are shown in Table 5. They were averaged over 10 noisy versions of each image, and in the last column, the standard deviation of these results is indicated for each LSAI.

From these results it can be observed that the orientation of the anisotropic LSAI is very important. Except for the case where $\sigma=5$, the denoising result is better when the anisotropic LSAI is oriented in the direction of lowest correlation compared to when the LSAI is oriented in the direction of highest correlation. The difference in terms of PSNR depends on the image and increases with increasing standard deviation of the added noise. The explanation for this trend lies in the fact that AWGN is transformed into correlated noise by the curvelet transform. When calculating the LSAI of a coefficient in a neighborhood that coincides with the direction of this correlation, it will be contaminated by the same noise that disturbed the coefficient and therefore will not be a good indicator of the local spatial activity, even though the significant coefficients are highly correlated along this direction [see Table 3 (a)]. This is increasingly true for higher noise levels (cf the rising trend of the correlation between insignificant coefficients and neighbors in the direction of highest correlation in Fig. 4). Because significant coefficients are still somewhat correlated along the direction of lowest correlation [see Table 3(b)], calculating the LSAI in a neighborhood oriented in this direction will lead to a better denoising performance. For very small standard deviations of the noise ($\sigma=5$), the advantage of high correlation among significant coefficients will dominate the undesirable influence of the correlated noise along the direction of highest correlation. For such small noise levels, calculating the LSAI in a neighborhood oriented in this direction leads to better denoising results.

Table 5 ProbShrinkCurv denoising results of some 512×512 grayscale images using the different LSAIs of Sec. 3.3. The noisy input images were contaminated with AWGN with different standard deviations σ. This table shows the denoising results in terms of PSNR using (a) a 1×5 anisotropic LSAI oriented in the direction of highest correlation, (b) a 1×5 anisotropic LSAI oriented in the direction of lowest correlation, (c) an adjacent, opposing, and parent LSAI, and (d) a combined IIB LSAI. The last column shows the estimated standard deviation of the results for each LSAI.

		PSNR (dB)				
LSAI	σ	Lean	Barbaba	Peppers	Baboon	σ_{PSNR}
(a) Intra-band anisotropic LSAI corr.	5	38.61	37.16	36.56	34.69	0.013
(b) Intra-band anisotropic LSAI uncorr.		38.53	37.14	36.45	34.59	0.015
(c) AOP LSAI		**38.80**	**37.27**	**36.70**	**34.92**	0.013
(d) Combined IIB LSAI		38.64	37.16	36.37	34.63	0.014
(a) Intra-band anisotropic LSAI corr.	10	35.36	33.62	33.97	30.41	0.023
(b) Intra-band anisotropic LSAI uncorr.		35.65	33.84	34.12	30.38	0.023
(c) AOP LSAI		35.71	33.71	34.20	**30.59**	0.022
(d) Combined IIB LSAI		**35.85**	**33.89**	**34.22**	30.48	0.022
(a) Intra-band anisotropic LSAI corr.	20	31.54	29.90	30.94	26.58	0.026
(b) Intra-band anisotropic LSAI uncorr.		32.24	30.36	31.34	26.60	0.026
(c) AOP LSAI		32.34	30.15	31.49	**26.82**	0.025
(d) Combined IIB LSAI		**32.57**	**30.45**	**31.63**	26.76	0.018
(a) Intra-band anisotropic LSAI corr.	30	29.08	27.68	28.69	24.61	0.046
(b) Intra-band anisotropic LSAI uncorr.		30.08	28.29	29.37	24.73	0.044
(c) AOP LSAI		30.18	28.02	29.55	**24.92**	0.043
(d) Combined IIB LSAI		**30.44**	**28.37**	**29.73**	24.87	0.043
(a) Intra-band anisotropic LSAI corr.	50	26.19	24.79	25.62	22.39	0.059
(b) Intra-band anisotropic LSAI uncorr.		27.38	**25.49**	26.53	22.66	0.056
(c) AOP LSAI		27.45	25.02	26.69	**22.78**	0.053
(d) Combined IIB LSAI		27.70	25.45	**26.81**	22.75	0.054

Table 5 further shows that the AOP LSAI outperforms the intra-band LSAI in the direction of lowest correlation for all the tested images except for *Barbara*. This result confirms our observations in Sec. 3.3. Indeed, the correlation of the intra-band LSAI with significant coefficients is smaller than in the AOP LSAI case. *Barbara* is the exception here because the correlation of the significant coefficients with the intra-band LSAI in the direction of lowest correlation is extremely high, much higher than the correlation with the AOP LSAI, and much higher than for the other tested images. The importance of this intra-band correlation for *Barbara* explains why the LSAI exploiting this correlation performs better in the denoising method.

For standard deviations larger than 5, the denoising performance of the combined intra- and inter-band LSAI is superior to that of the AOP LSAI. This is somewhat surprising, because we have not observed a greater average correlation between this LSAI and the significant coefficients for the images of our test set (see Sec. 3.3). A possible explanation is that the intra- and inter-band LSAIs contribute complementary information to the denoiser. The LSAI contributes to the determination of the level of significance of each coefficient to be denoised. When adding intra- to inter-band information, individual coefficients that correlate well with the inter-band LSAI but not with the intra-band LSAI will be judged as less significant, but others will behave in the opposite way. In other words, the total fraction of significance over all the coefficients in the sub-bands is not increased, but spread over more coefficients. This leads to a better denoising result for the tested

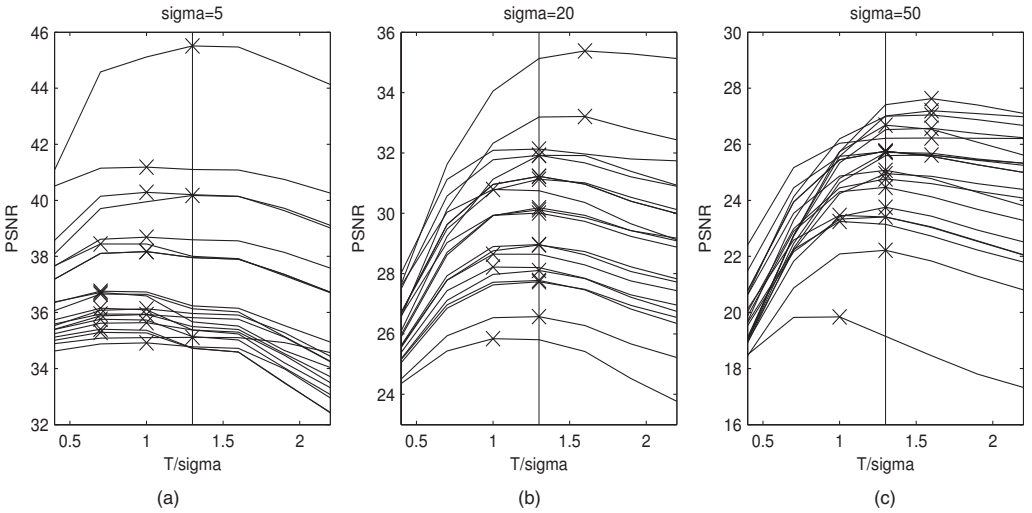

Fig. 6 Denoising performance of the ProbShrinkCurv method as a function of the ratio threshold to noise standard deviation T/σ for several images of the USC-SIPI image test set and for noise levels (a) $\sigma=5$, (b) $\sigma=20$, and (c) $\sigma=50$. The optimal choice of the ratio T/σ for each image is marked by a cross. The solid vertical line marks $T=1.3\sigma$.

images (except for a small deterioration for *Baboon*). For $\sigma=5$, the information contributed by the uncorrelated intra-band LSAI to the combined intra- and inter-band LSAI deteriorates the denoising performance. We had already observed that the uncorrelated intra-band LSAI performs poorly for such low noise levels.

Because of the superior denoising performance of the IIB LSAI for all noise levels except the very low ones, we will use it in our ProbShrinkCurv method to compare with state-of-the-art denoising methods (see Sec. 6).

5 Choice of the Threshold *T*

A crucial issue that has not been addressed up to this point is the choice of the threshold T. This threshold determines our actual signal of interest. This signal of interest should be chosen to minimize the MSE of the denoised image. An analytical derivation seems intractable for the assumed prior. Nonetheless, to make this choice in a theoretically founded way, we follow the approach of Pižurica *et al.*[22] and Jansen *et al.*[37] and base ourselves on the "oracle" thresholding described by Mallat.[38] Oracle thresholding provides us with the MMSE estimate of transform coefficients corrupted with AWGN by zeroing the ones with noise-free components below the standard deviation σ of the noise. Thus, $T=\sigma$ marks the boundary between significant and insignificant coefficients.

Direct application of this estimator to our denoising method is unrealistic for several reasons. First, this approach requires an oracle to inform us of the value of the noise-free coefficient in order to make our classification decision about the noise-contaminated coefficient. Such an oracle is not available in a realistic scenario. Second, this choice minimizes the MSE when denoising is achieved by hard thresholding the noisy coefficients. Our denoiser shrinks rather than thresholds the coefficients. Finally, the noisy curvelet coefficients that we consider are not contaminated with white but with colored noise. Considering all these factors, we expect the optimal value of the threshold T not to coincide exactly with the standard deviation of the contaminating noise in the sub-bands, but to peak in its vicinity.

To verify this expectation, we show in Fig. 6 the influence of the threshold T on the denoising performance of ProbShrinkCurv for several noise levels and for the test images of the Miscellaneous volume of USC-SIPI (the images used in Sec. 6 were removed from the test set to avoid overfitting of T). For each image, results were calculated for seven different values of T, each time averaged over 10 noisy versions, and for noise standard deviations $\sigma=5$, 20, and 50. Figure 6 plots the resulting curves for half of the images (not for all to avoid overloading the graphs). The optimal choice of the ratio T/σ for each image is marked by a cross. These figures show that the optimal value for T does indeed always lie in the vicinity of $T=\sigma$, but also that it is image and noise-level dependent. The overall trend is that at lower noise levels, the best denoising performance is achieved for lower values of T.

The average quality drop over all the images (except for the images used in Sec. 6) when fixing the ratio T/σ to a particular value is quantified in Table 6 for noise levels $\sigma=5$, 10, 20, 30, and 50. We can observe first that at moderate noise levels, the average quality drop when choosing a value for T/σ within 23% of the optimum does not exceed 0.5 dB. Second, we can observe that the trend of lower noise levels favoring a lower threshold and vice versa is confirmed. However, from $\sigma=10$ onwards, the smallest overall drop in performance is achieved when keeping T/σ constant, namely at 1.3. From these experiments, we chose $T=1.3\sigma$ throughout this paper (also for the statistical analysis of Sec. 3).

6 Results

In this section, we report on the denoising performance of our newly developed method and provide a comparison with some state-of-the-art denoisers.

Table 6 Average denoising quality drop (in dB) over all the images in the USC-SIPI test set when fixing the ratio threshold T to standard deviation σ to a particular value.

σ	$T/\sigma=0.4$	$T/\sigma=0.7$	$T/\sigma=1.0$	$T/\sigma=1.3$	$T/\sigma=1.6$	$T/\sigma=1.9$
5	0.7591	0.0714	**0.0374**	0.3707	0.4502	1.1447
10	1.9985	0.6963	0.1253	**0.1089**	0.3646	1.0022
20	3.8806	1.4579	0.3252	**0.0754**	0.2777	0.7480
30	4.6897	2.2665	0.4825	**0.0761**	0.2404	0.6448
50	5.4500	2.4833	0.7104	**0.1028**	0.2328	0.5495

6.1 *ProbShrinkCurv Denoising Results*

Denoising results of the ProbShrinkCurv method on some 512×512 and 256×256 grayscale images are reported in Table 7. When possible, we used the versions of the images included with the online implementation of Ref. 20. In these experiments, the standard deviation of the AWGN was assumed to be known. Results were averaged over 10 noise realizations for each image and for each noise level. The standard deviation of the results for each noise level are reported in the last column of Table 7. We used 4 scales in the curvelet decomposition and 16 orientations at the coarsest level. Varying these parameters alters the denoising results. The optimal numbers are image dependent, but we found that this choice produces satisfying results for a broad class of images.

6.2 *Numerical Comparison with Other Denoisers*

Figure 7 compares the results of the newly developed method to some state-of-the-art denoisers. These include BiShrink, which uses a dual tree complex wavelet decomposition;[16] BLS-GSM with the parameters set as in Ref. 20 and operating on a full steerable pyramid decomposition of the image; the BM3D method as reported in Ref. 39; and the ProbShrink method for wavelets in its redundant wavelet transform implementation.[22] A comparison to simple curvelet domain hard thresholding is also provided (4 scales in the curvelet decomposition, 16 orientations at the coarsest level, and the threshold at $k\sigma$ with $k=4$ at the finest scale and 3 otherwise). Implementations of curvelet hard thresholding and of the methods in Refs. 16, 20, 22, and 39 are publicly available and were used to produce the results of Fig. 7. Results were averaged over 10 noise realizations for each image and for each noise level. For all methods, we assumed that the standard deviation of the noise was known for the denoising technique.

These results show that the ProbShrink method adapted to curvelets outperformed or matched its wavelet-based counterpart for all images. Improvements were smallest for *Lena* and *Baboon* and greatest for *Barbara* and *Peppers* (up to 1.08 dB). Differences were more pronounced for large standard deviations of the AWGN. In fact, for standard deviations $\sigma=5$ and smaller, the ProbShrink method for wavelets performed better. These results were not included in Fig. 7 to avoid overloading it. This trend complies with

Table 7 Denoising results in terms of PSNR (dB) of some 512×512 and 256×256 grayscale images. The last column shows the estimated standard deviation of the results for each noise level.

σ/PSNR	Lena 512×512	Barbara 512×512	Boats 512×512	Baboon 512×512	House 256×256	σ_{PSNR}
2/42.03	42.38	42.58	42.35	42.00	43.49	0.014
5/34.13	37.86	37.16	36.19	34.63	38.26	0.015
10/28.13	35.20	33.86	33.12	30.48	34.84	0.033
15/24.61	33.38	31.83	31.23	28.15	32.92	0.049
20/22.13	32.02	30.38	29.93	26.76	31.50	0.032
25/20.23	30.97	29.23	28.88	25.63	30.41	0.042
30/18.70	30.01	28.29	28.01	24.87	29.45	0.057
35/17.44	29.21	27.43	27.23	24.15	28.59	0.123
50/14.61	27.14	25.32	25.35	22.75	26.60	0.102

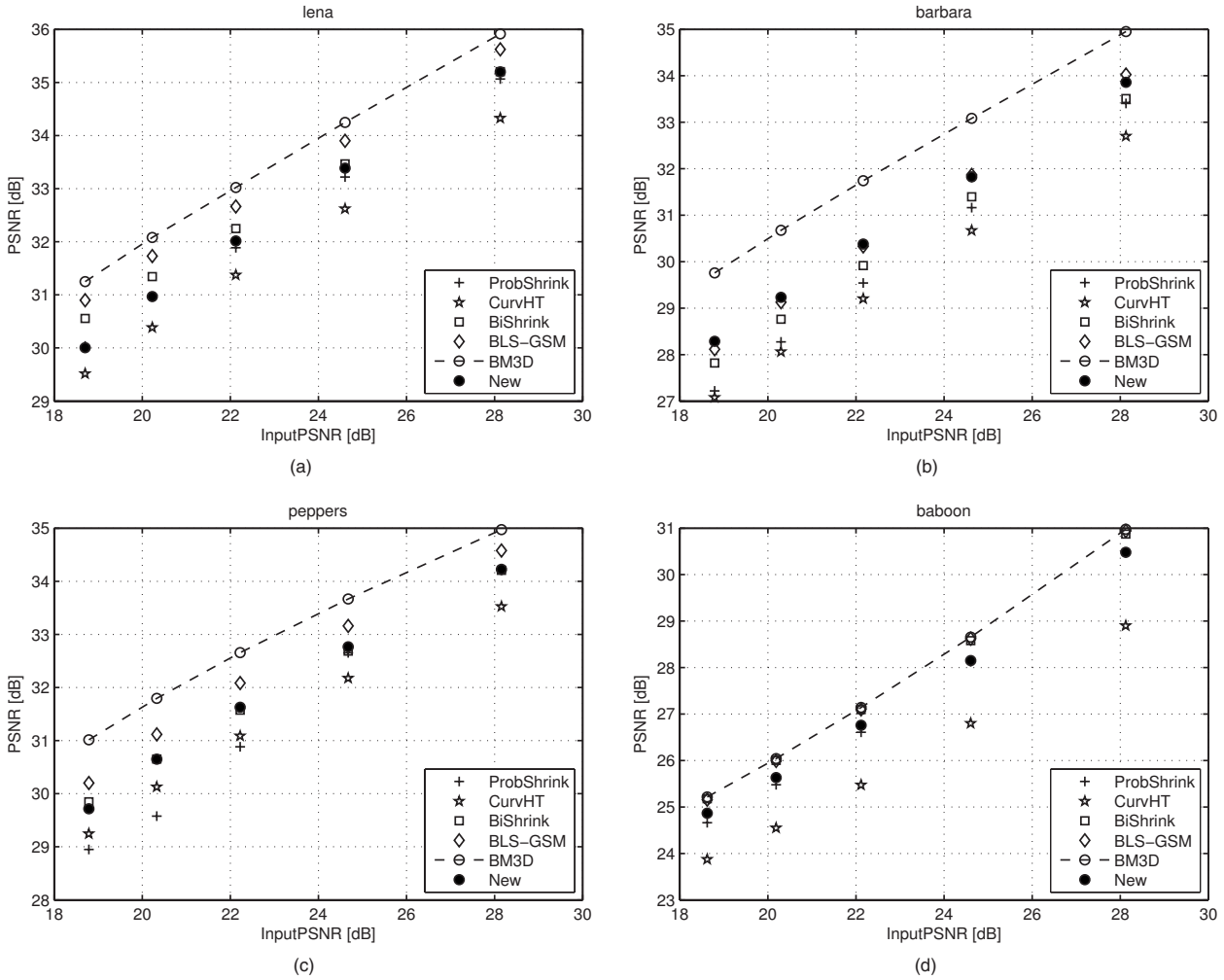

Fig. 7 Output PSNR as a function of input PSNR for four 512×512 images: (a) *Lena*, (b) *Barbara*, (c) *Peppers*, and (d) *Baboon*. The following six methods are shown: the ProbShrink method for wavelets in its redundant wavelet transform implementation,[22] curvelet hard thresholding (Curv HT), BiShrink using a dual tree complex wavelet decomposition,[16] BLS-GSM with the parameters set as in Ref. 20, the BM3D method as reported in Ref. 39, and the proposed ProbShrinkCurv (New) method.

our observations from Secs. 4.5 and 5 in which we found that for small noise levels, the AOP LSAI and a smaller threshold T were more appropriate than the IIB LSAI with $T=1.3\sigma$, which we chose because of the other noise levels.

The improvements of ProbShrinkCurv over simple curvelet domain hard thresholding were considerable for all images at all noise levels, but they were most notable for *Barbara* and *Baboon*, i.e., for images that were not sparsely represented in the curvelet domain. For these images, the more complex Bayesian and neighborhood-adaptive approach of ProbShrinkCurv provided a clear advantage over simple hard thresholding.

The performance of ProbShrinkCurv compared to other state-of-the-art techniques was somewhat image and noise-level dependent, but overall our new denoiser was competitive with BiShrink (based on a transform of similar redundancy as ProbShrinkCurv, which was about 7.2 for our choice of parameters), but it was outperformed by BM3D for all images and by BLS-GSM for all images except *Barbara*, for which denoising results were similar.

6.3 *Visual Comparison with Other Denoisers*

Figures 8 and 9 visually compare some cut-outs of the denoising results. The good edge-preserving qualities of ProbShrinkCurv are illustrated on both the feathers in *Lena*'s hat and the striped pattern on *Barbara*'s scarf, which are sharply preserved. Compared to ProbShrink for wavelets and BiShrink, the method is less plagued by over-smoothing and impulse-like artifacts. However, some stripe-like artifacts are visible, although less than in the curvelet hard thresholding case. The denoising results of ProbShrinkCurv also give a much sharper impression than the results obtained through hard thresholding.

6.4 *Denoising with Unknown Noise Variance*

If the standard deviation of the noise is not known, as is often the case in practical situations, one must estimate it from the corrupted data, e.g., using the mean absolute deviation estimator of Donoho *et al.*[40] The inaccuracy of this estimate affects the denoising performance of the methods.

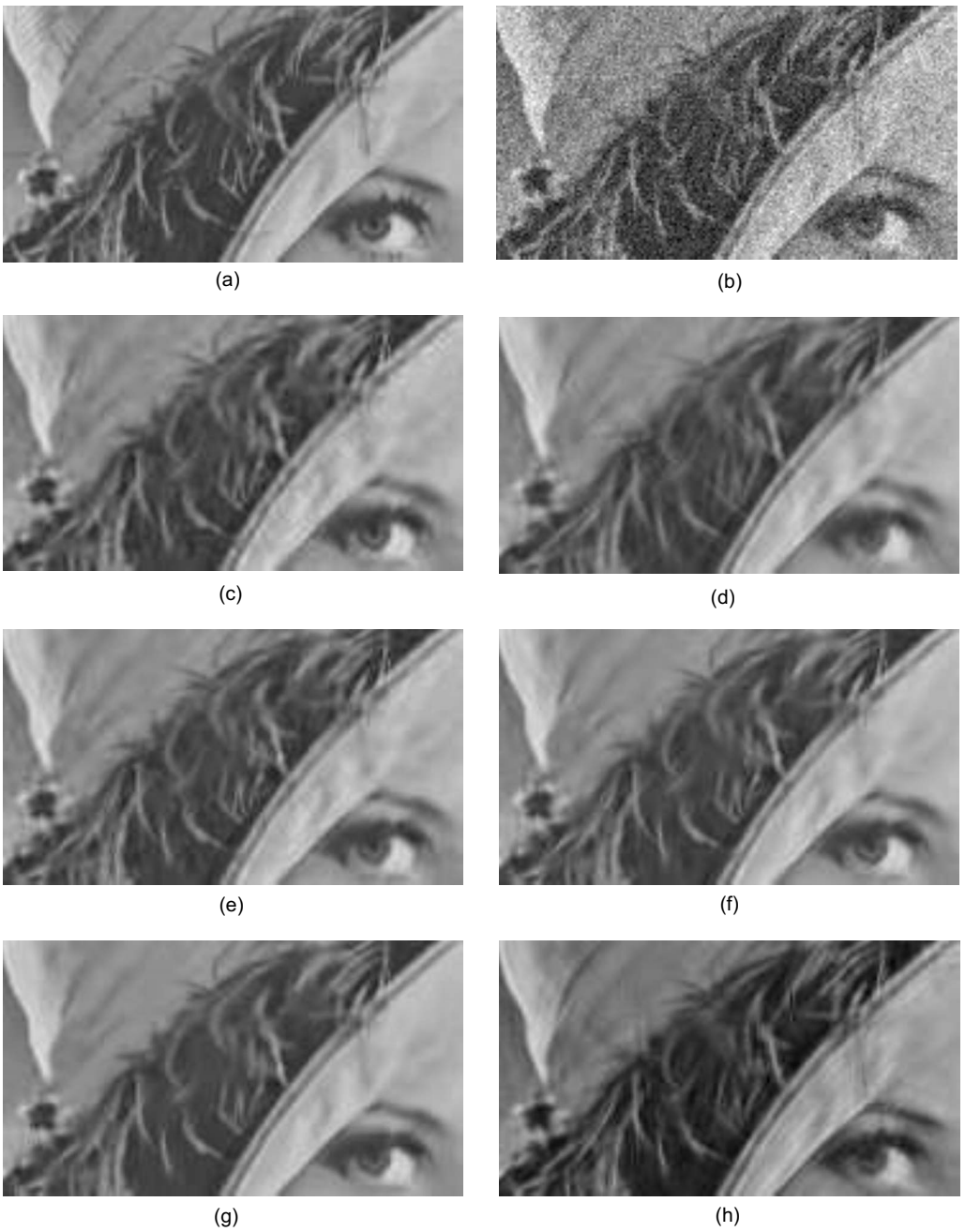

Fig. 8 Details of image *Lena*: (a) original image, (b) noisy image (noise standard deviation 20), plus the denoising results of (c) ProbShrink, (d) curvelet hard thresholding, (e) BiShrink, (f) BLS-GSM, (g) BM3D, and (h) ProbShrinkCurv.

Table 8 shows the difference in PSNR performance between denoising with known and estimated noise variances for several methods and noise levels, averaged each time over 10 noise realizations of the images *Lena*, *Barbara*, *Peppers*, and *Baboon*. In the curvelet-based methods, the noise variance was estimated from the last orientation sub-band at the finest scale and in the BiShrink method from the first finest-scale sub-band in the dual tree complex wavelet decomposition (as it was implemented in the BiShrink code available online). In the other denoisers, no noise estimation was implemented, so they were not included in Table 8. From this table it is obvious that the denoising performance of the compared methods drops dramatically for low noise levels but is more or less robust to inaccuracies in noise variance estimation at higher noise levels. We can conclude that the performance differences are of the same order of magnitude for the three methods compared here.

6.5 *Execution Times*

Table 9 compares the mean execution times of BiShrink, BLS-GSM, BM3D, ProbShrink, curvelet hard thresholding, and ProbShrinkCurv when denoising a 512×512 grayscale

Fig. 9 Details of image *Barbara*: (a) original image, (b) noisy image (noise standard deviation 20), plus the denoising results of (c) ProbShrink, (d) curvelet hard thresholding, (e) BiShrink, (f) BLS-GSM, (g) BM3D, and (h) ProbShrinkCurv.

image. Results were averaged over 10 experiments on a AMD Athlon(TM) 64 3400+2.40-GHz processor. All algorithms were run in Matlab and were not optimized for speed. The mean execution time of ProbShrinkCurv was lower than the mean execution time of the methods with the best denoising performance (BLS-GSM and BM3D) and was also lower than that of its wavelet-based counterpart. The method is about half as fast as curvelet hard thresholding.

7 Conclusion

In this paper, we showed the differences in statistical behavior between curvelet coefficients containing a significant noise-free component and those in which no signal of interest was present. We also discussed the adaptation of the ProbShrink denoising method for wavelets[22] to curvelets, resulting in a method that we call ProbShrinkCurv. In particular, we put the knowledge gained from our statistical study to use in the design of an appropriate LSAI for this new method.

When considering intra-band coefficients for the LSAI, we found that, although curvelet coefficients were more correlated along the principal direction of their generating basis function, neighboring coefficients in the perpendicular direction were a better indicator of the significance of the reference coefficient in terms of denoising results. We further ascertained that it is beneficial to also incorporate coefficients from adjacent, opposing, and parent sub-bands in the LSAI.

The resulting denoising method, ProbShrinkCurv, outperformed its wavelet-based counterpart and produced re-

Table 8 PSNR difference (in dB) between denoising with known and with estimated noise standard deviation σ averaged over the images *Lena*, *Barbara*, *Peppers*, and *Baboon*.

Denoiser	$\sigma=2$	$\sigma=10$	$\sigma=20$	$\sigma=30$	$\sigma=50$
CurvHT	4.36	0.38	0.08	−0.01	−0.08
BiShrink[16]	4.14	0.25	0.09	0.08	0.04
ProbShrinkCurv	4.41	0.24	0.06	0.04	0.11

Table 9 Mean execution times when denoising a 512×512 gray-scale image. Results are averaged over 10 experiments on a AMD Athlon(TM) 64 3400+2.40-GHz processor.

Denoising method	Mean execution time (s)
BiShrink[16]	2.14
BLS-GSM[20]	75.19
BM3D[39]	11.98
ProbShrink[22]	10.52
Curvelet hard thresholding	4.93
ProbShrinkCurv	8.87

sults that were both visually competitive with and numerically close to those of state-of-the-art denoisers.

Acknowledgments

The authors would like to thank the reviewers for their many helpful comments and suggestions that greatly improved this manuscript. Author L. Tessens is supported as a research assistant by the research foundation Flanders (FWO Flanders). A. Pižurica is a postdoctoral research fellow of FWO Flanders.

References

1. E. J. Candès and D. L. Donoho, "Ridgelets: a key to higher-dimensional intermittency," *Philos. Trans. R. Soc. London, Ser. A* **357**(1760), 2495–2509 (1999).
2. D. Donoho, "Wedgelets: nearly minimax estimation of edges," *Ann. Stat.* **27**, 859–897 (1999).
3. M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.* **14**(12), 2091–2106 (2005).
4. E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.* **5**(3), 861–899 (2006).
5. E. Candès, "The curvelet transform for image denoising," *IEEE Inter. Conf. Image Process.*, vol. 1, p. 7 (Thessaloniki) (2001).
6. J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.* **11**(6), 670–684 (2002).
7. D. D.-Y. Po and M. N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Process.* **15**(6), 1610–1620 (2006).
8. A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.* **15**(10), 3089–3101 (2006).
9. D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Image Process.* **41**(3), 613–627 (1995).
10. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.* **9**(9), 1532–1546 (2000).
11. F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a bayesian approach," *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **60**, 725–749 (1998).
12. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Signal denoising using adaptive bayesian wavelet shrinkage," *Proc. IEEE-SP Inter. Symp. on Time-Frequency and Time-Scale Analysis*, 225–228 (1996).
13. M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika* **85**(2), 391–401 (1998).
14. E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," *IEEE Inter. Conf. Image Process.*, 1, 379–382 (1996).
15. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors," *IEEE Trans. Inf. Theory* **45**(3), 909–919 (1999).
16. L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.* **9**(12), 438–441 (2002).
17. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.* **9**(9), 1522–1531 (2000).
18. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *Proc. SPIE* **3816**, 31–44 (1999).
19. M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.* **6**(12), 300–303 (1999).
20. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Process.* **12**(11), 1338–1351 (2003).
21. J. A. Guerrero-Colon and J. Portilla, "Two-level adaptive denoising using gaussian scale mixtures in overcomplete oriented pyramids," *Proceedings Inter. Conf. Image Process. (ICIP)*, 1, Genova, Italy, 105–108 (2005).
22. A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.* **15**(3), 654–665 (2006).
23. L. Boubchir and J. M. Fadili, "Multivariate statistical modeling of images with the curvelet transform," *Proc. 8th Inter. Symp. Signal Process. and its Applications (ISSPA 2005)*, 2, Sydney, Australia, 747–750 (2005).
24. L. Boubchir and J. M. Fadili, "Modélisation statistique multivariée des images dans le domaine de la transformée de curvelet," *Proc. 20th GRETSI Symp. on Signal and Image Proc.*, Louvain-la-Neuve, Belgique 233–236 (2005).
25. A. Alecu, A. Munteanu, A. Pizurica, W. Philips, J. Cornells, and P. Schelkens, "Information-theoretic analysis of dependencies between curvelet coefficients," in *Proc. ICIP* (2006).
26. F. Abramovich, P. Besbeas, and T. Sapatinas, "Empirical bayes approach to block wavelet function estimation," *Comput. Stat. Data Anal.* **39**(4), 435–451 (2002).
27. F. Abramovich and T. Sapatinas, "Bayesian approach to wavelet decomposition and shrinkage," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds., Lecture Notes in Statistics 141, pp. 33–50, Springer-Verlag, New York 1999).
28. B. Vidakovic and F. Ruggeri, "Bams method: Theory and simulations," *Sankhya* **63**(2), 234–249 (2001).
29. B. Vidakovic, "Nonlinear wavelet shrinkage with bayes rules and bayes factors," *J. Am. Stat. Assoc.* **93**, 173–179 (1998).
30. L. Tessens, A. Pižurica, A. Alecu, A. Munteanu, and W. Philips, "Spatially adaptive image denoising based on joint image statistics in the curvelet domain," *Proc. SPIE* **6383**, 63830L (2006).
31. L. Tessens, A. Pižurica, A. Alecu, A. Munteanu, and W. Philips, "Modeling curvelet domain inter-band image statistics with application to spatially adaptive image denoising," *Proc. ProRISC - Program for Research on Integrated Systems and Circuits, Workshop 2006*, 208–213, Veldhoven, The Netherlands, November (2006).
32. E. Candès and D. Donoho, "Curvelets—a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surfaces*, L. L. Schumaker *et al.*, Ed., Vanderbilt University Press, Nashville (1999).
33. I. M. Johnstone and B. W. Silverman, "Empirical bayes selection of wavelet thresholds," *Ann. Stat.* **33**(4), 1700–1752 (2005).
34. USC-SIPI image database, http://sipi.usc.edu/services/database/, last accessed 19 September 2008.
35. R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-28**(2), 137–145 (1980).
36. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-32**(6), 1109–1121 (1984).
37. M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. Am. Stat. Assoc.* **96**, 629–639 (2001).
38. S. Mallat, *A Wavelet Tour of Signal Processing*, ch. X, sec. 10.2, Academic Press (1998).
39. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
40. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika* **81**(3), 425–455 (1994).

Linda Tessens received her MS in electrical engineering in 2005 at Ghent University, Belgium. Since 2005 she has been working at the Department of Telecommunications and Information Processing of Ghent University in the Image Processing and Interpretation research group as a research assistant for the Flemish Fund for Scientific Research (FWO). Her main research interests are image analysis and vision networks.

Aleksandra Pižurica received the Diploma degree in electrical engineering in 1994 from the University of Novi Sad (Serbia and Montenegro), the MSc diploma in telecommunications in 1997 from the University of Belgrade (Serbia and Montenegro) and the PhD degree in applied sciences in 2002 from Ghent University (Belgium). Currently she is a postdoctoral researcher at the Department of Telecommunications and Information Processing of Ghent University. Her research interests include image restoration, multi-resolution representations, Markov Random Field models, signal detection and estimation, and optical communication systems.

Alin Alecu received a MSc in electrical engineering and computer science from Politehnica University of Bucharest (UPB) in 1998, and the PhD degree in applied sciences from Vrije Universiteit Brussels (VUB), Brussels, Belgium, in 2004. From 1998 to 1999, he was a teaching assistant at UPB. From 1999 until 2007 he was a member of the Department of Electronics and Informatics (ETRO) of VUB. Since 2008 he has been with Oracle Corporation, where he holds the position of Principle Engineer of R&D, leading the R&D group of Oracle Romania. His research interests include Bayesian networks, data mining, and knowledge management. He is the author of more than 20 scientific publications.

Adrian Munteanu received a MSc in electronics and telecommunications in 1994 from the Politehnica University of Bucharest, Romania, the MSc degree in biomedical engineering in 1996 from the Technical University of Patras, Greece, and the PhD degree in applied sciences in 2003 from Vrije Universiteit Brussels (VUB), Brussels, Belgium. Since October 1996, he has been with the Department of Electronics and Informatics (ETRO) of VUB, and since 2006 he has held a professorship at ETRO. His research interests include scalable still-image and video coding, multiresolution image analysis, image and video transmission over networks, video segmentation and indexing, scalable mesh coding, error resilient coding, and statistical modeling. He is the author or co-author of more than 150 scientific publications, patent applications, and contributions to standards, and has contributed to four books in his areas of interest.

Wilfried Philips received a Diploma degree in electrical engineering in 1989 and the PhD degree in applied sciences in 1993, both from Ghent University, Belgium. From October 1989 until October 1997 he worked at the Department of Electronics and Information Systems of Ghent University for the Flemish Fund for Scientific Research (FWO-Vlaanderen), first as a research assistant and later as a postdoctoral research fellow. Since November 1997 he has been with the Department of Telecommunications and Information Processing of Ghent University, where he is currently a full-time professor and heads the Image Processing and Interpretation research group, which has recently become part of the virtual Flemish ICT research institute IBBT. Some of the recent research activities in the group include image and video restoration and analysis, and the modeling of image reproduction systems. Important application areas targeted by the group include remote sensing, surveillance, and industrial inspection.